

Proof of Concept for the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A)

Release Webinar

Rachel Marks & Alli Coritz

Population Division, U.S. Census Bureau

January 31, 2023



The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. DRB Approval Number: CBDRB-FY22-DSEP-005, CBDRB-FY23-0158

Data Protection in the 2020 Census Products

- Census Bureau has historically tabulated data for detailed race and ethnic groups and American Indian and Alaska Native tribes and villages while protecting information provided by respondents.
- Over time, more and more granular census data have been published.
- Advances in data science, more powerful computers, and externally accessible ‘big data’ has increased the risk of identifying individuals from published statistics.
- To mitigate, the Census Bureau is implementing new disclosure avoidance methods for the 2020 Census.
- The 2020 Census Disclosure Avoidance System (DAS) is based on the framework of differential privacy, which adds noise – or variations from the actual count – to the collected data.
- Census Bureau’s goal is to meet the data needs of data users while protecting respondents’ confidentiality.
- The feedback we receive from you is critical as we finalize the settings for the Detailed DHC-A.

2020 Census Data Products – Race and Ethnicity

Released

Upcoming Releases

**Redistricting File
(Public Law 94-171)**
*Released on
August 12, 2021 and
September 16, 2021*

This product provides population counts and data on the voting age population for major race and ethnic groups.

**Demographic Profile
and
Demographic and Housing
Characteristics File (DHC)**
Planned Release: May 2023

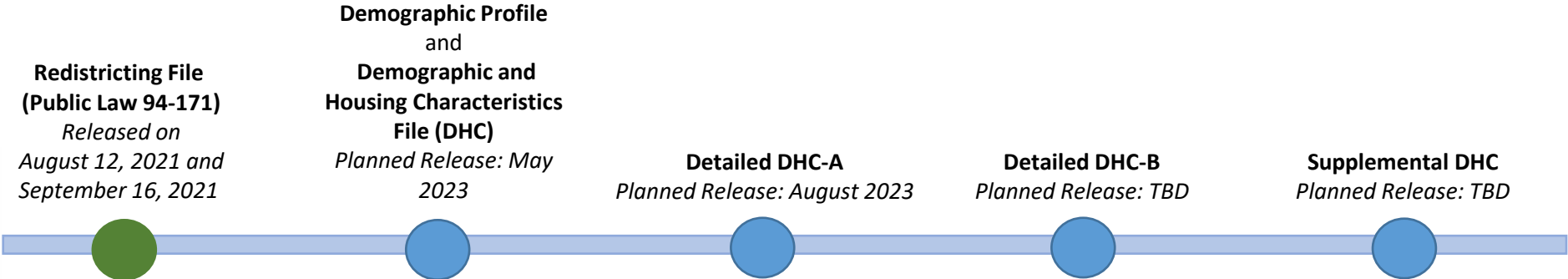
Supplemental DHC (S-DHC)
Planned Release: TBD

These products provide demographic and housing characteristics for the major race and ethnic groups.

Detailed DHC-A
Planned Release: August 2023

Detailed DHC-B
Planned Release: TBD

The Detailed DHC-A and Detailed DHC-B are the successors to the Summary File 2 and the American Indian and Alaska Native Summary Files. They provide data for disaggregated race and ethnic groups and AIAN tribes and villages.



What Data are Included in the Detailed DHC-A?

- Subjects repeated by approximately 370 detailed racial and ethnic groups and 1,200 detailed American Indian and Alaska Native (AIAN) tribes and villages:
 - Total population
 - Sex by age for selected age categories
- Geographic levels included:
 - Nation
 - State
 - County
 - Tracts
 - Places
 - American Indian/Alaska Native/Hawaiian Home Land (AIANNH) areas
- Uses an adaptive design that allows us to provide more detailed statistics for racial and ethnic groups with larger populations, while at the same time ensuring sufficient confidentiality protections.

What Data are Included in the Detailed DHC-A?

- **Data are included for *detailed* and *regional* race and Hispanic origin groups and AIAN tribes and villages**
 - Detailed groups include disaggregated groups such as Puerto Rican, Chinese, the Navajo Nation, Samoan, etc.
 - Regional groups include groups such as Caribbean, East Asian, American Indian, Polynesian, etc.
- **Data are included for detailed and regional *race alone* and *race alone or in any combination groups***
 - Race alone groups include people who reported a single entry (e.g., Korean) and no other detailed race group.
 - Race alone or in any combination groups include people who reported a single entry (e.g., Korean) and people who reported that entry with one or more other race(s) (e.g., Korean *and* Thai, or Korean *and* Native Hawaiian). This represents the maximum number of people who reported as that detailed race group, either alone or in any combination with one or more additional detailed race group(s).

What Does the Proof of Concept Include?

- Reflects the proposed differential privacy settings for the 2020 Detailed DHC-A (i.e., privacy loss budget, margins of error, population thresholds, and levels of geography).
- Applies differential privacy, using the SafeTab-P algorithm, to 2010 Census data for tables that will be included in the 2020 Detailed DHC-A.
 - Note: Data users should be mindful when analyzing comparisons. Differential privacy is applied to the non-private 2010 data, whereas the published 2010 data has undergone “swapping” (another type of disclosure avoidance). Differences between the two types of counts therefore reflect the two different privacy methods used.
- The product includes technical documentation, summary metrics, and examples of tabulated data for the public to review and provide feedback on.

What is NOT Included in the Proof of Concept?

- This product does not include summary files or a complete 2010 Census dataset.
 - Instead, we provide a subset of tabulated data for users to review.
- Because this product was created using 2010 Census data, it does not reflect the improvements made to the 2020 Census race question and code list.
- Therefore, this product does not include the full range of detailed and regional groups that will be available in the final 2020 Census Detailed DHC-A.
 - The improvements to the race question and code list will be reflected in the final 2020 Census Detailed DHC-A released in August.

Differences Between the TopDown and SafeTab-P Differentially Private Algorithms

TopDown (Redistricting Data (P.L. 94-171), DHC, Demographic Profile)

- Algorithm produces privacy-protected microdata
- All geographies aggregate as expected
- When aggregating data, the statistical noise generally cancels out and the statistics become *more* accurate
- Consistency across data products
- Overall accuracy can be targeted, but the exact levels of accuracy cannot be known in advance (*except in the case of the Redistricting Data (P.L. 94-171) Summary File*)
- Does not use adaptive design

SafeTab-P (Detailed DHC-A)

- Algorithm directly produces privacy-protected tabulations
- There is no requirement that geographies aggregate as expected
- When aggregating data, it generally becomes more variable the more you aggregate
- Not consistent with other 2020 Census data products
- All margins of error are determined in advance and met 95% of the time
- Uses adaptive design to determine the amount of data provided

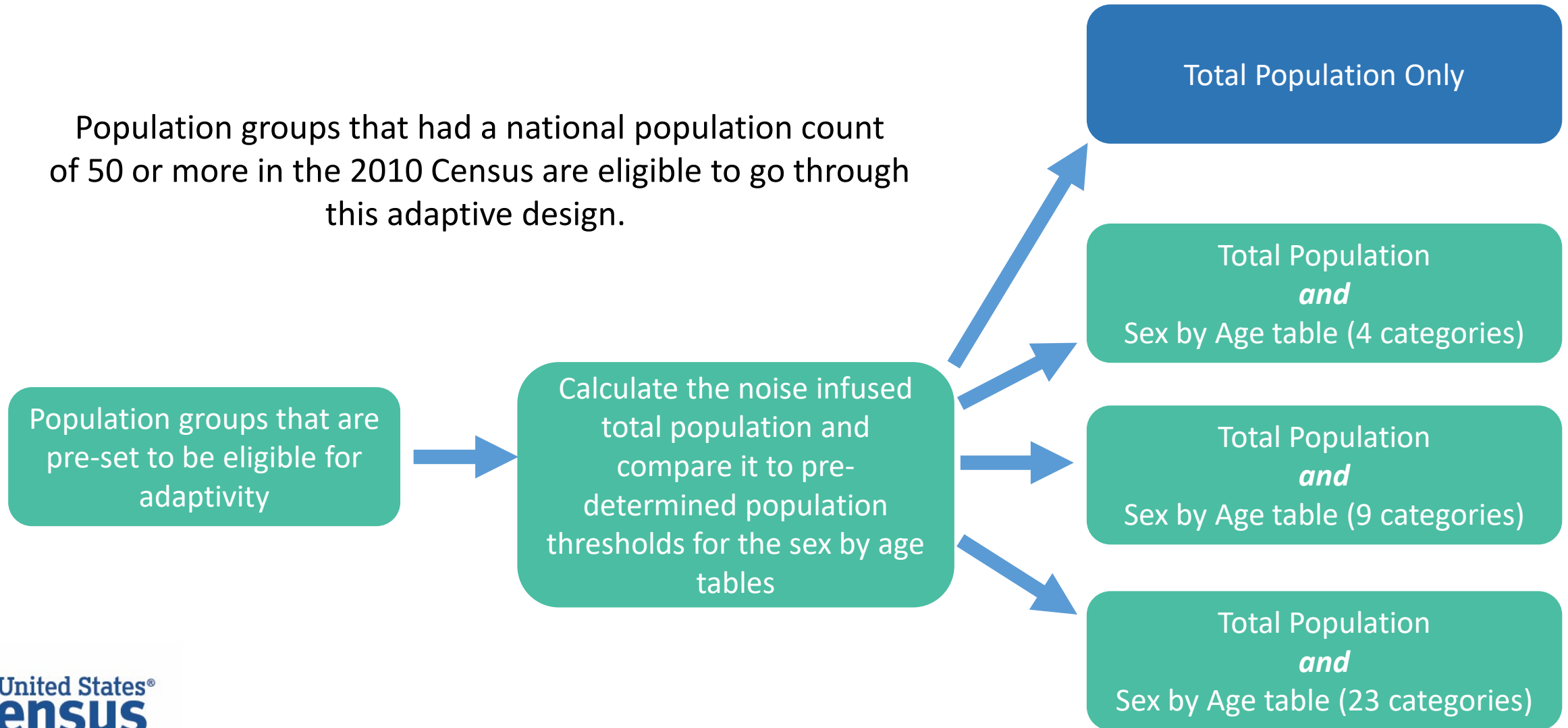
Using Adaptive Design to Produce the Detailed DHC-A



Detailed groups with a national population count less than 50 in the 2010 Census are pre-set to receive total population only in the Detailed DHC-A.

Using Adaptive Design to Produce the Detailed DHC-A

Population groups that had a national population count of 50 or more in the 2010 Census are eligible to go through this adaptive design.



Sex by Age 23 categories

Sex x Age(23)		
Age Group	Male	Female
Under 5 years	146	147
5 to 9 years	130	131
10 to 14 years	109	107
15 to 17 years	79	77
18 and 19 years	68	97
20 years	34	89
21 years	64	76
22 to 24 years	150	151
25 to 29 years	235	265
30 to 34 years	237	302
35 to 39 years	230	374
40 to 44 years	177	356
45 to 49 years	187	372
50 to 54 years	139	275
55 to 59 years	84	151
60 and 61 years	27	53
62 to 64 years	30	60
65 and 66 years	7	15
67 to 69 years	12	27
70 to 74 years	12	25
75 to 79 years	13	12
80 to 84 years	3	7
85 years and over	2	3

Sex by Age 9 categories

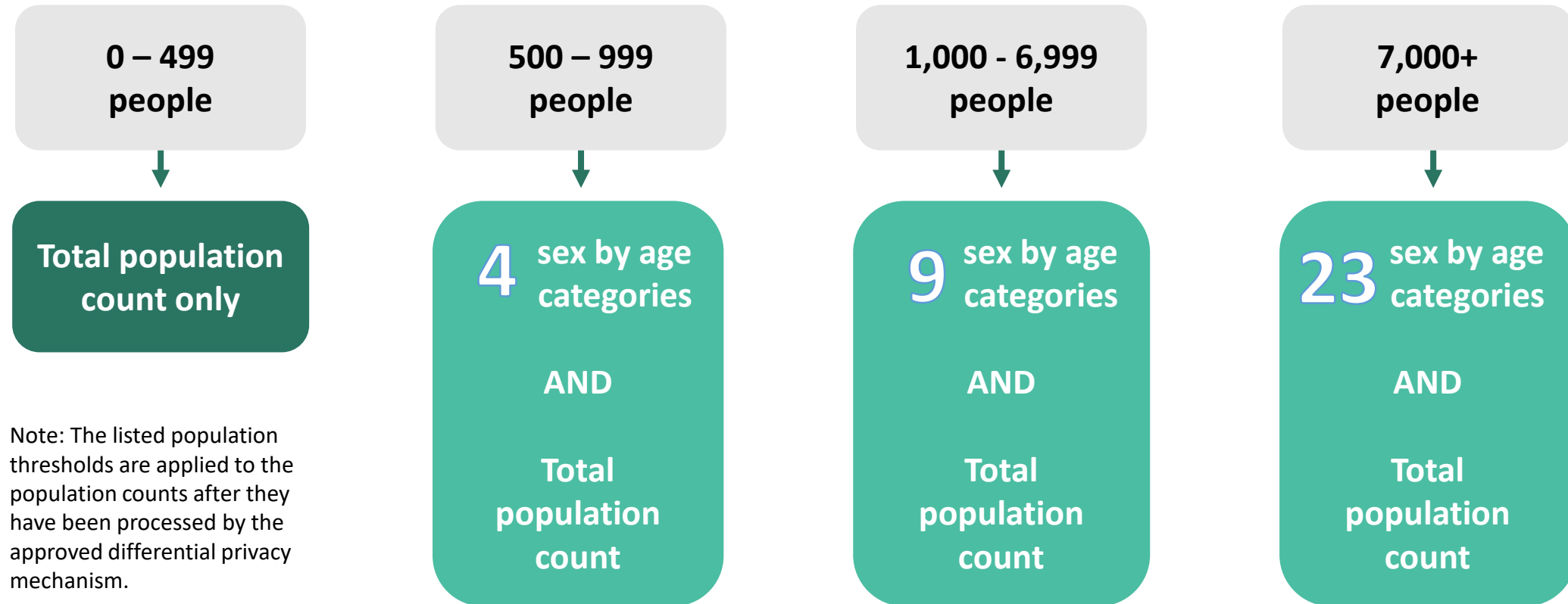
Sex x Age(9)		
Age Group	Male	Female
Under 5 years	146	147
5 to 17 years	318	315
18 to 24 years	316	413
25 to 34 years	472	567
35 to 44 years	407	730
45 to 54 years	326	647
55 to 64 years	141	264
65 to 74 years	31	67
75 years and over	18	22

Sex by Age 4 categories

Sex x Age(4)		
Age Group	Male	Female
Under 18 years	464	462
18 to 44 years	1,195	1,710
45 to 64 years	467	911
65 years and over	49	85

Minimum Population Counts (Thresholds) for Detailed Groups

Nation and State Levels

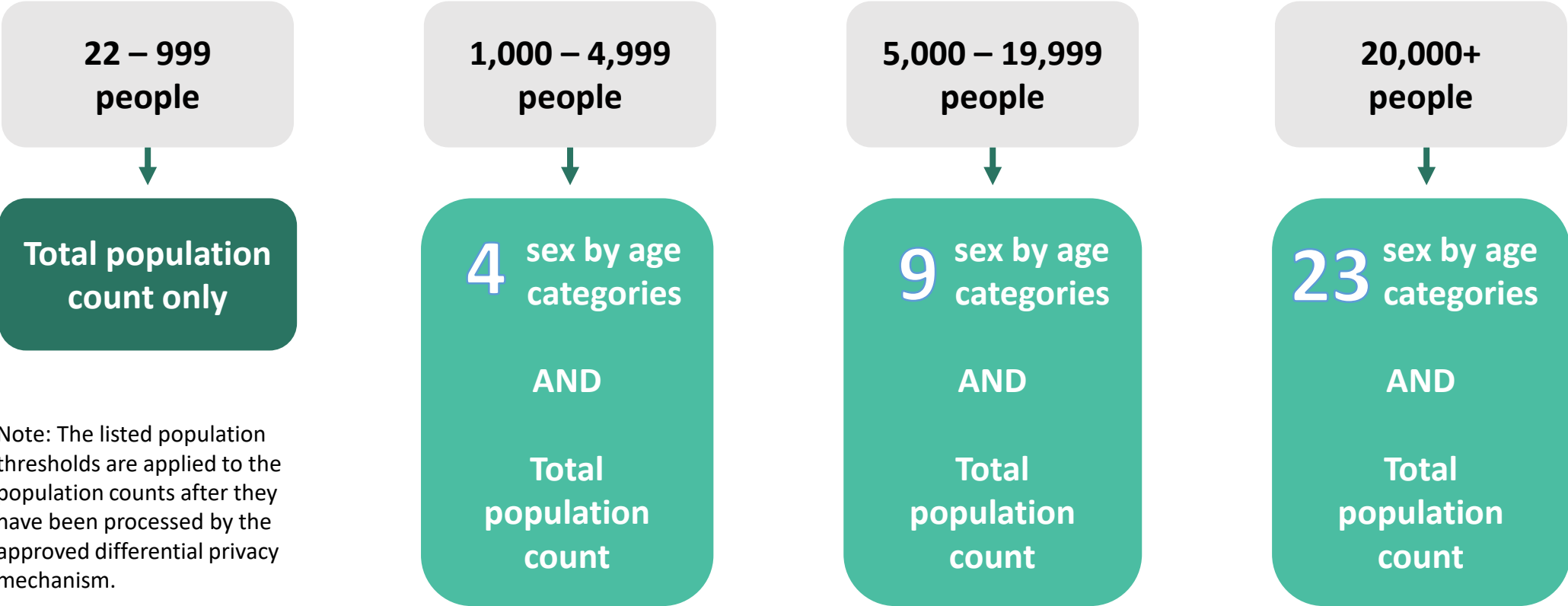


Note: The listed population thresholds are applied to the population counts after they have been processed by the approved differential privacy mechanism.

Margin of Error = 3 for detailed groups at nation and state levels

Minimum Population Counts (Thresholds) for Detailed Groups

County, Tract, Place, and AIANNH Areas

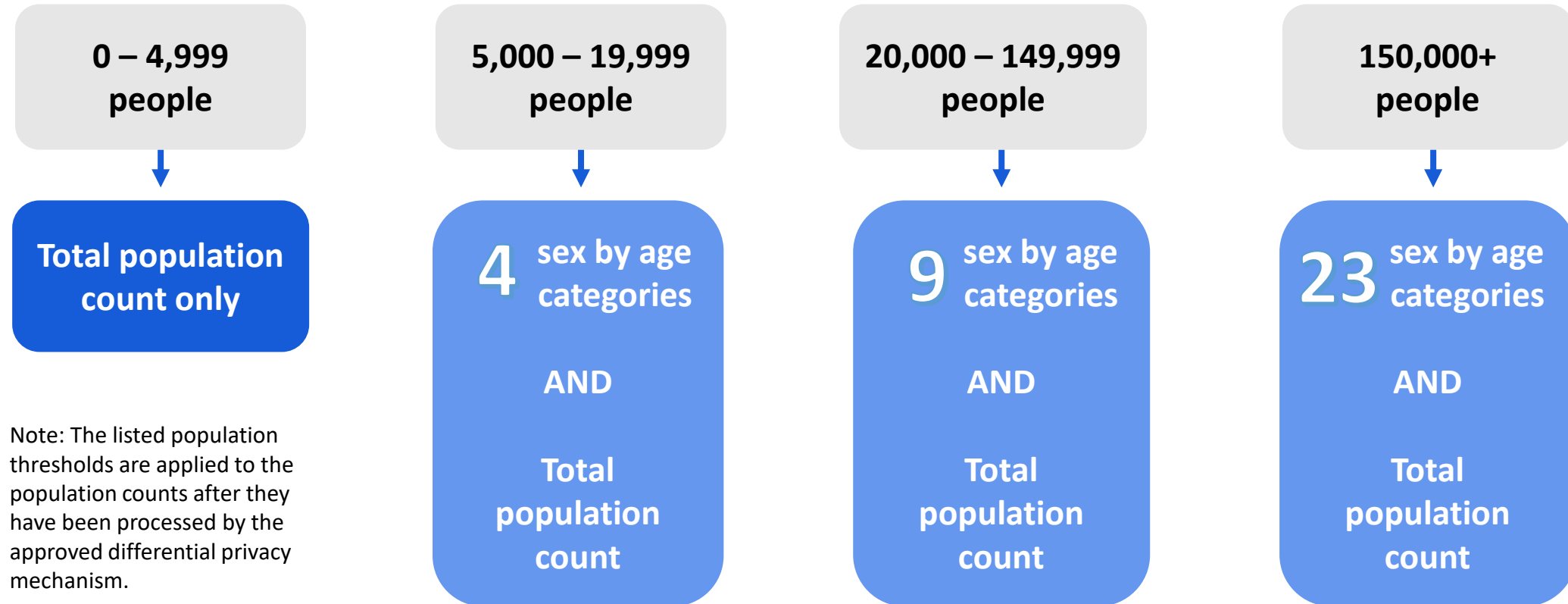


Note: The listed population thresholds are applied to the population counts after they have been processed by the approved differential privacy mechanism.

Margin of Error = 11 for detailed groups at county, tract, place, and AIANNH area levels

Minimum Population Counts (Thresholds) for Regional Groups

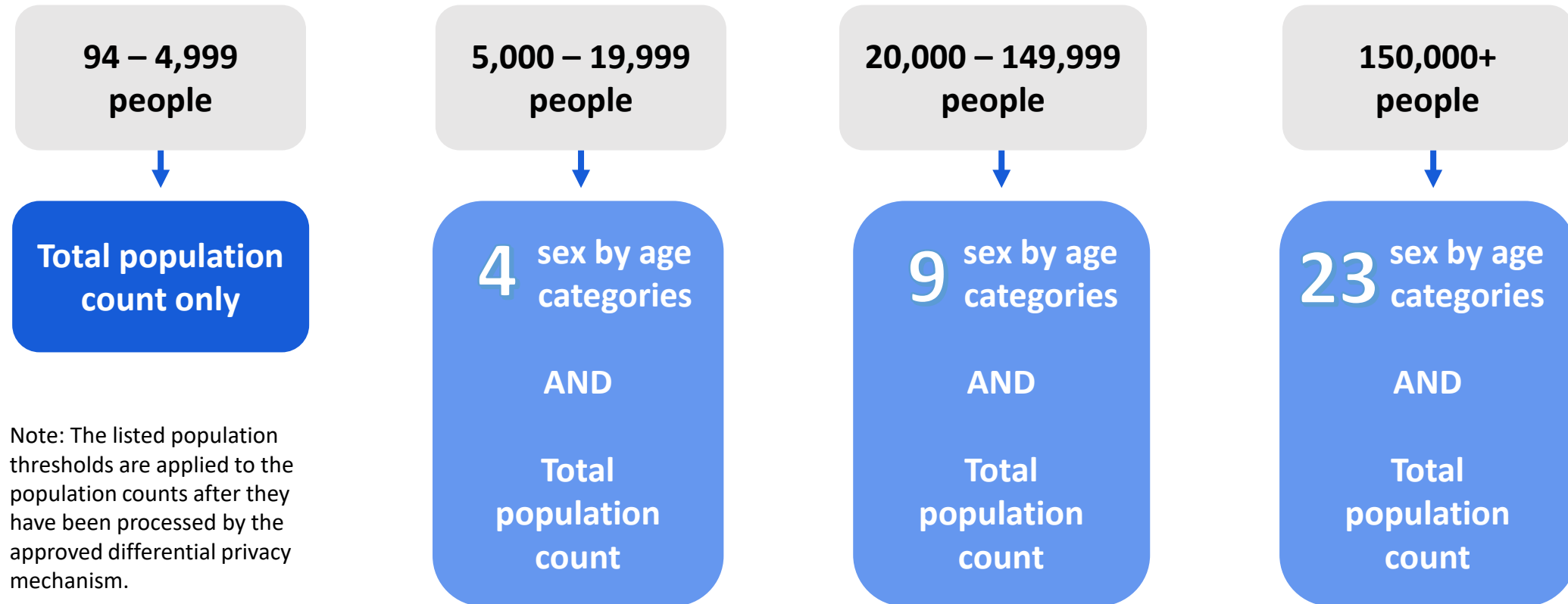
Nation and State Levels



Margin of Error = 50 for regional groups at nation and state levels

Minimum Population Counts (Thresholds) for Regional Groups

County, Tracts, and Places



Note: The listed population thresholds are applied to the population counts after they have been processed by the approved differential privacy mechanism.

Margin of Error = 50 for regional groups at county, tract, and place levels

Summary of Results

- At the national level:
 - Over 95% of the noise infused counts for each racial and ethnic category are within their expected MOE.
 - The average difference between the enumerated count and the noise infused count for racial and ethnic groups and American Indian and Alaska Native tribes and villages at the national level ranges from +/- 0.5 to +/-1.35.
 - Less than 1% of noise infused counts are outliers for each racial and ethnic category, meaning they are outside their MOE by more than twice the value of the MOE.
 - For example, given an MOE of 3, fewer than 1% of noise infused counts are +/-6 from their enumerated counts.

You can explore more details about the accuracy of these data in the Summary Metrics.

Limitations of the Data

- There are two sources of improbable data in the file:
 - The noise necessary to protect respondent confidentiality.
 - The lack of internal consistency requirements compared to other 2020 Census data products, such as the Redistricting File and DHC.
- As a result, data users may observe inconsistencies, including some implausible results, in the Detailed DHC-A.
- These inconsistencies help preserve random noise and unbiased data in the product, as well as preserving the MOEs of the noise infused data with 95% confidence.
- Data users may encounter unavailable data because of suppression implemented on implausible data and/or populations with small counts at lower levels of geography.

For more information about limitations of the data, please see the Proof of Concept on [census.gov](https://www.census.gov).

Guidance for Data Users

- Data users should always use the published count for the specific combination of race/ethnic group and geographic area they are interested in.
- We do not recommend creating custom aggregations, as this will result in less accurate data.

Guidance for Data Users

- Detailed groups may not sum to their corresponding regional group.
- Data users should use the published count for the regional group rather than creating custom regional groups.

Universe: Caribbean regional group			
United States	Caribbean regional group total	Sum of detailed Caribbean groups (<i>Cuban + Dominican + Puerto Rican + Other Caribbean</i>)	Difference between regional group total and sum of detailed groups
2010 Noise infused count	7,828,451	7,828,335	116

Guidance for Data Users

- Values from lower-level geographies may not sum to values from higher-level geographies.
- Data users should use the published counts for the geography of interest rather than creating custom aggregations.

Universe: Native Hawaiian alone or in any combination			
	United States	Sum of states	Difference
2010 Published count	527,077	527,077	0
2010 Noise infused count	527,209	527,108	101

Guidance for Data Users

- Lower-level geographies may have larger counts than higher-level geographies.
- Data users should use the published counts for the geography of interest.

Universe: Village of Red Devil alone			
	United States	Alaska	Difference
2010 Published count	3	3	0
2010 Noise infused count	2	3	1

Public Feedback Request

- Review the Proof of Concept materials and assess the levels of accuracy, population thresholds, and differences between the protected data in this product and the published 2010 Census data.
- If we release the Detailed DHC-A using the settings in the Proof of Concept, will your work be impacted positively, negatively, or not at all?
- Email your feedback to 2020DAS@census.gov by Thursday, March 2, 2023.

We are seeking your feedback before the disclosure avoidance settings for the Detailed DHC-A are finalized.